基于多种深度学习算法的田间玉米籽粒检测与计数

刘晓航[1,2]，张　昭[1,2*]，刘嘉滢[1,2]，张　漫[1,2]，李　寒[1,2]，
Paulo FLORES[3]，韩雄哲[4,5]

（1. 中国农业大学智慧农业系统集成研究教育部重点实验室，北京 100083；2. 中国农业大学农业农村部农业信息获取技术重点实验室，北京 100083；3. 北达科他州州立大学农业与生物工程系，法戈 58102，美国；4. 韩国江原大学生物系统工程系，春川 24341，韩国；5. 韩国江原大学智慧农业交叉学科，春川 24341，韩国）

摘　要：为快速准确获取玉米收获过程中遗失籽粒数信息，进行收割损失调节等管理，对比评估了单阶段和两阶段主流目标检测网络对田间玉米籽粒计数的性能。首先，利用RGB相机获取包含不同背景和不同光照的图像数据，并进一步生成数据集；其次，构建籽粒识别的不同目标检测网络，包括Mask R-CNN、EfficientDet-D5、YOLOv5-L、YOLOX-L，并利用所采集的420幅有效图像对构建的四种网络进行训练、验证、测试，图像数分别为200、40和180幅；最后，依据测试集图像的识别结果进行籽粒计数性能评价。试验结果表明，YOLOv5-L网络对测试集图像检测的平均精度为78.3%，模型尺寸仅为89.3 MB；籽粒计数的检测正确率、漏检率和F_1值分别为90.7%、9.3%和91.1%，处理速度为55.55 f/s，识别与计数性能均优于Mask R-CNN、EfficientDet-D5和YOLOX-L网络，并对具有不同地表遮挡程度和籽粒聚集状态的图像具有较强的鲁棒性。深度学习目标检测网络YOLOv5-L可实现实际作业中玉米收获损失籽粒的实时监测，精度高、适用性强。

关键词：收获损失；田间玉米籽粒；深度学习；籽粒计数；YOLOv5-L；YOLOX-L；Mask R-CNN；EfficientDet-D5

中图分类号：S126　　**文献标志码：**A　　**文章编号：**SA202207004

引用格式：刘晓航，张昭，刘嘉滢，张漫，李寒，FLORES Paulo，韩雄哲. 基于多种深度学习算法的田间玉米籽粒检测与计数[J]. 智慧农业(中英文)，2022，4(4)：49-60.

LIU Xiaohang, ZHANG Zhao, LIU Jiaying, ZHANG Man, LI Han, FLORES Paulo, HAN Xiongzhe. Infield corn kernel detection and counting based on multiple deep learning networks[J]. Smart Agriculture, 2022, 4(4): 49-60. (in Chinese with English abstract)

1　引　言

玉米作为传统粮食作物之一，因其兼具"粮-经-饲"等多元属性，已成为世界上种植面积最广和贸易量最大的农作物[1]。相比于水稻和小麦，玉米机械收获过程中籽粒损失率相对较高，开展玉米田间收获籽粒损失监测研究，对评估联合收获机作业性能，实现损失自主调节，保障粮食实收产量等具有重大意义。

目前玉米收获损失主要是基于传感器进行检测，即通过安装在收割机不同位置（即清选筛或排草口）处的传感器（如光、声音、微波、压电陶

收稿日期：2022-07-12
基金项目：中央高校基本科研业务费专项资金资助（00112502）；世界顶尖涉农大学国际合作交流种子基金（15052001）
作者简介：刘晓航（1996—），男，博士研究生，研究方向为智慧农业。E-mail：lxhhaust@163.com
*通信作者：张　昭（1985—），男，博士，教授，研究方向为智慧农业。E-mail：zhaozhangcau@cau.edu.cn

瓷、压电薄膜等）来获取籽粒冲击感应板的频率、幅值等信号特征，并借助信号处理、机器学习等方法构建损失籽粒预测模型，以实现田间玉米收获损失率的实时监测[2,3]。但受籽粒冲击角度、速度差异等因素的影响，实际检测中易造成误识别，且受敏感材料自身特性（安装位置、灵敏度、有效区域范围等）、作物状态以及收割机作业工况，如喂入量、草谷比、行进速度、滚筒转速等因素的影响，在依据单一环节（夹带和清选等）的监测估计最终的收获损失时，其结果可靠性不高。因此，受各项复杂因素联合影响的传感器检测方法难以满足实际作业中对玉米收获损失率监测的精准、高效需求，亟需一种能够直接、快速对玉米收获时损失籽粒进行精准计数的方法。

机器视觉技术应用于玉米籽粒品质分级、质量估计以及损伤检测的研究验证了其用于籽粒识别的可行性[4-6]。基于此，研究者提出了利用图像处理进行谷物收获损失检测的方法[7-9]，即通过灰度化、去噪、分割等方法检测谷物籽粒，并结合对谷物籽粒形状、颜色、面积等属性的分析完成收获籽粒的损失计数。但受阈值不一致及图像自身差异的影响，籽粒计数结果的可靠性和稳定性不高，并仍以收获过程中某一环节（脱粒、清选等）的损失检测为主，忽略了对直接反映收获损失的田间地表籽粒的检测。随着深度学习目标检测技术的进步，其在提高物体检测精度、效率、鲁棒性等方面表现出了极大潜力[10]。为此，Monhollen等[11]提出了采用目标检测网络直接识别田间地表籽粒的损失评估方法，并开发了相应的图像采集平台，达到了82%的损失检测精度。但在图像采集前，需对残余物进行清除以显露玉米籽粒，操作繁琐，且计数精度受残余物清除程度的影响。由上述分析可知，现有检测方法的精度、适用性等与实际作业的理想需求仍有一定差距，而深度学习的使用使得玉米收获损失监测精度有了进一步提升的可能。

因此，本研究的主要目的是评估深度学习技术直接进行真实地表籽粒计数的可行性与表现，以简化收获损失检测步骤，实现监测精度与适用性的综合提升。本研究的主要工作包括：（1）利用RGB相机采集玉米收获后的真实地表图像数据；（2）构建两阶段目标检测网络掩膜区域卷积神经网络（Mask Regions Convolutional Neural Network，Mask R-CNN）和单阶段目标检测网络EfficientDet-D5、YOLOv5-L、YOLOX-L进行玉米籽粒计数；（3）分析不同地表遮挡程度和籽粒聚集状态等场景及网络类别对最终计数性能的影响，并总结出用于田间玉米收获损失籽粒计数的最优深度学习模型。

2 材料与方法

2.1 图像采集

供试数据采集于美国北达科他州大福克斯县（North Dakota，Grand Forks，US）的玉米试验田，拍摄对象为收获时遗留地表的玉米籽粒。为避免收割过程中秸秆烟尘影响图像采集质量，在联合收获机（John Deere X Series，Moline，IL，US）完成玉米收获后，由研究人员手持佳能EOS Rebel T7i型相机（图像分辨率2000×2000，帧率6 f/s，自动曝光、自动对焦模式）在距地面约1.3 m的垂直高度对收割之后的地面进行拍摄，田间收获场景与图像采集设备如图1所示。样本采集时间为2020年11月7日上午8:00—11:00，共获取500幅图像。

2.2 技术路线

本研究旨在使用深度学习算法对采集图像进行分析，实现对田间玉米籽粒的自动检测与计数。计数方法流程如图2所示，主要包含3个环节：

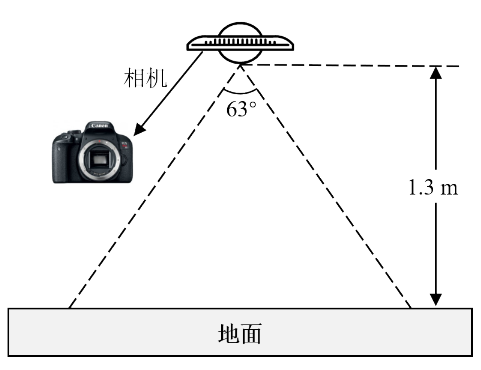
（1）构建数据集：筛选采集图像有效帧，对其进行场景分类和籽粒标注，以构建玉米籽粒目标检测数据集。

（2）玉米籽粒计数：构建并训练用于籽粒目标检测的不同网络模型，以实现田间玉米收获损失籽粒的实时计数。

（3）结果分析：对可视化模型训练过程及结



(a)田间玉米收获场景

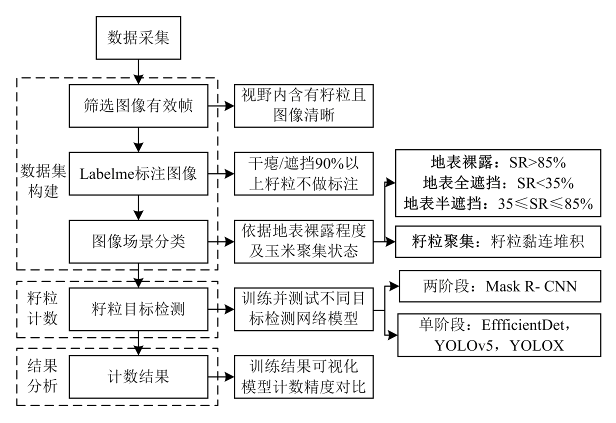


(b)地面图像采集

图1 田间玉米收获场景与地面图像采集示意图

Fig. 1 Infiled corn harvest combine in working mode and schematic for image acquisition

果进行分析，在测试集上对不同模型方法进行评估，并推荐最优模型。



注:SR是指地表像素面积占比

图2 田间玉米籽粒计数的整体技术路线图

Fig. 2 General technical route for corn kernel counting

2.2.1 数据集构建

为保证采集图像有效性以便模型训练与测试，通过人工筛选出含有玉米籽粒且视野清晰的420幅图像（包含6773个玉米籽粒）作为总数据集。随机选取200幅图像（包含1628个玉米籽粒）作为训练集，其余40幅（包含224个玉米籽粒）和180幅（包含4921个玉米籽粒）分别作为验证集和测试集，其中验证集用以调节训练模型的超参数，避免过拟合。使用Labelme软件进行数据标注，并按照COCO（Common Objects in COntext）格式构建目标检测数据集。最后为准确评估网络模型的适用性及其对收获减损的指导性能，依据田间秸秆对地表的遮挡程度和玉米籽粒聚集状态，将180幅测试集图像划分为4类（图3）：地表裸露（60幅，1415个玉米籽粒）、地表半遮挡（60幅，1372个玉米籽粒）、地表全遮挡（31幅，218个籽粒）和籽粒聚集（29幅，1916个籽粒）。其中，地表裸露、地表半遮挡和地表全遮挡分别指场景内地表像素与图像总像素比值在（0.85, 1）、［0.35, 0.85］和（0, 0.35）区间的图像，且玉米籽粒之间相互离散；籽粒聚集通常是指相互黏连和堆叠的籽粒数大于12个的图像。场景命名中的遮挡是指秸秆对地表的遮挡。

2.2.2 方法设计

为避免传统目标检测算法存在的识别准确率低、模型适用性差、特征依赖性强等问题[12, 13]，选取优势明显且应用广泛的深度学习网络对获取的图像进行检测，以简化特征设计、区域选择的繁琐过程，减少人工构建特征的有效性、鲁棒性对检测精度与效率的影响，实现玉米籽粒的高精度实时检测。由于基于深度学习的目标检测算法可分为基于区域建议的两阶段方法和基于回归分析的单阶段方法。因此为对比两类方法在籽粒检测任务中的适用性，分别从两类方法中选取网络。对于两阶段方法而言，现有的R-CNN、空间金字塔池化网络（Spatial Pyramid Pooling Network, SPP-Net）、Fast R-CNN、Faster R-CNN等网络的性能均低于具备目标检测与分割双重功能的Mask R-CNN网络[14-16]，且不考虑目标分割效果时，Mask R-CNN与Faster-RCNN的功能一致[16]，故选用Mask R-CNN网络作为单阶段方法

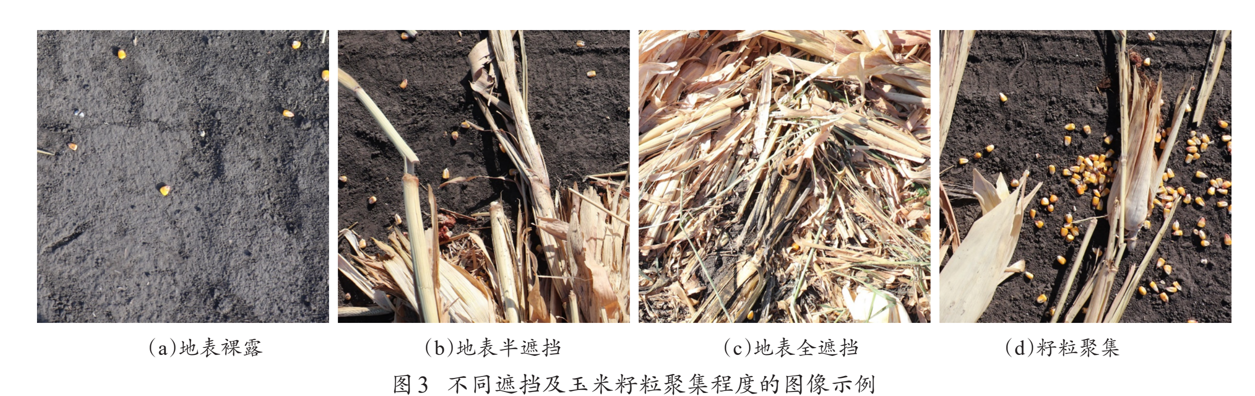


图3 不同遮挡及玉米籽粒聚集程度的图像示例

Fig.3 Image samples with different occlusion and corn kernel aggregation degree

的对照网络。而单阶段方法作为目标检测算法的主流方向，研究者已通过改进特征提取网络、多尺度融合机制、标签分配策略以及NMS-Free检测器等提出了不同的网络模型，并以YOLO系列最为经典和高效，故依据模型参数量、平均精度等指标从该系列网络中选取了在标准图像集COCO下检测性能较优的YOLOv5和YOLOX网络来对比有无锚框时的模型泛化性。此外还选取了EfficientDet网络来评估在固定资源限制下通过对网络深度、宽度、分辨率进行统一缩放实现玉米籽粒检测精度与效率平衡的可行性。

Mask R-CNN是在Faster R-CNN基础上，通过引入并行掩模分支和ROI Align算法来消除取整误差，实现精度提升的两段式检测网络[17]。因能实现物体掩模的像素级输出，常被作为衡量其他网络优劣性能的标准，其结构如图4所示。图像输入后，主干网络生成并融合具有不同尺度和语义信息的特征图，区域建议网络（Region Proposal Network，RPN）和ROI Align层则负责查找特征图中最佳的目标候选区域并解决特征图谱与对应原图的错位问题，最后通过分类预测实现对候选区域内目标类别、位置及掩模的预测。

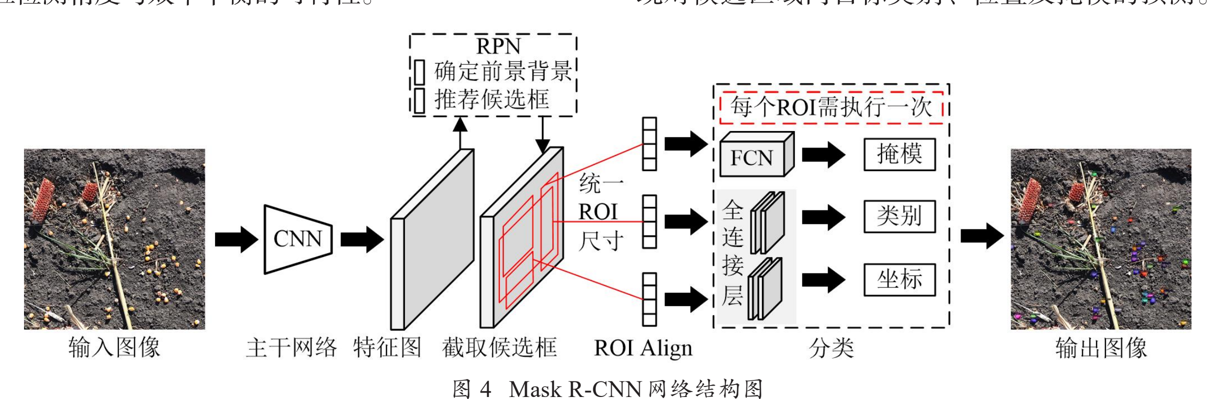


图4 Mask R-CNN网络结构图

Fig. 4 Structure of Mask R-CNN network

EfficientDet是单阶段目标检测模型，其依据神经结构搜索特征金字塔网络（Neural Architecture Search Feature Pyramid Network，NAS-FPN）的神经搜索和路径聚合网络（Path Aggregation Network，PANet）的双向融合思想，提出了能够快速实现多尺度特征融合的加权双向特征金字塔网络（Bidirectional Feature Pyramid Network，BiFPN），从而实现了检测精度与效率的显著提升[18]。如图5所示，EfficientDet结构主要由分别负责多尺度特征提取、融合以及目标位置与类别预测的主干特征提取网络、加强特征提取网络和预测网络组成。

YOLOv5继承了YOLOv4的模型架构，将输入图像划分为S×S个网格，并由目标中心所在网格预测其边界框位置、分类概率及置信度，是

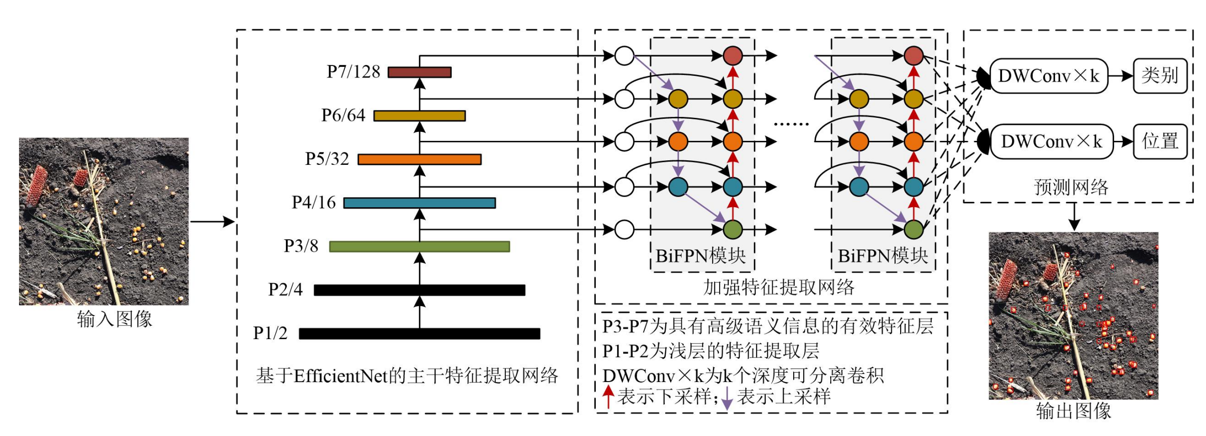


图 5 EfficientDet 网络结构

Fig. 5 Structure of EfficientDet network

目前表现较优的目标检测网络[19]，其结构如图6所示。图像数据经输入端预处理后，进入以跨阶段局部网络（Cross Stage Partial Network，CSPNet）和Focus为架构的主干网络中进行三种尺度的籽粒特征提取，并送入以特征金字塔网络（Feature Pyramid Network，FPN）和路径聚合网络结构为主的Neck模块中进行特征聚合，最后进入Head模块在特征图上使用锚定框得到带有置信度和框坐标信息的输出结果。为增强对遮挡、重叠目标籽粒的检测性能，试验中采用GIOU_Loss作为边界框的损失函数，并使用加权非极大值抑制算法对其进行筛选。

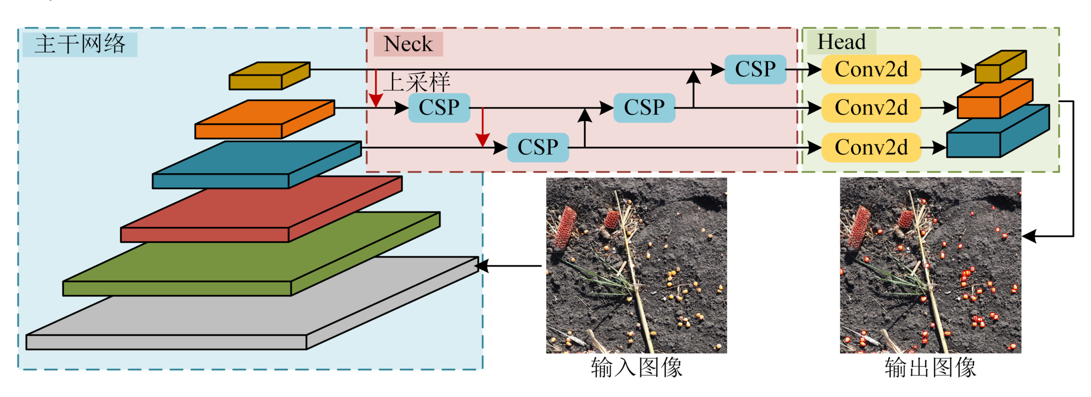


图 6 YOLOv5 网络结构

Fig. 6 Structure of YOLOv5 network

YOLOX是在总结当前深度学习领域多项研究成果和训练技巧的基础上对YOLOv3进行优化得到的网络[20]，其结构如图7所示。模型内部主干网络和Neck模块延续了YOLOv3的Dark-net53+SPP和FPN架构，并对输入端和Head模块进行了改进。即在输入端采用Mosaic和Mixup数据增强策略来丰富检测物体的背景；在Head模块引入Decoupled head、Anchor-free、Multi positives和SimOTA等操作来提高模型收敛速度与计算效率，并缓解正负样本的不平衡问题，实现全局信息下的最优样本匹配。此外Decoupled head中添加的交并比（Intersection over Union，IoU）分支能够有效增强预测能力，使得预测框的回归速度更快、精度更高。

由于所选四种网络中，EfficientDet、YOLOv5、YOLOX均有不同的模型系列，为在固定资源限制下最大化模型检测精度与效率，试验中选取EfficientDet-D5、YOLOv5-L和YOLOX-L作为三种网络的检测模型。网络训练的初始化参数设置为：初始学习率0.003、最大迭代次数100

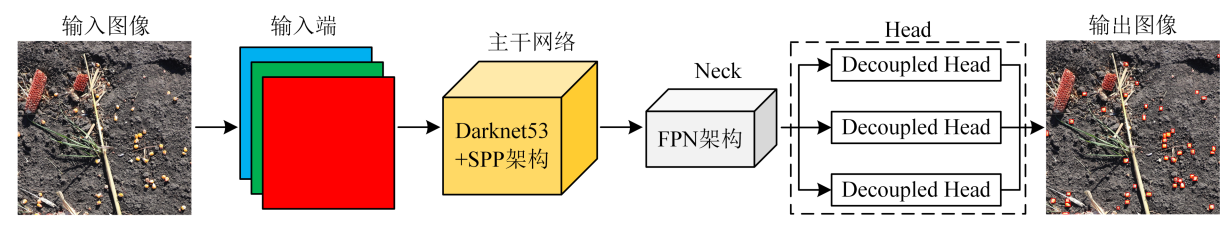


图7　YOLOX网络结构图

Fig. 7 Structure of YOLOX network

次、动量0.9、batchsize为4，优化器为Adam。由于Mask R-CNN为两阶段网络需要更多的迭代次数来实现收敛，依据经验其迭代次数设置为60,000次。

2.2.3　评价指标

为量化分析网络模型性能，本研究从模型复杂度、检测效果以及在移动端的应用潜力出发，采用目标检测的标准化评价指标衡量其在测试集的表现，主要包括平均精度（Average Precision，AP）、模型参数量、每秒10亿次的浮点运算次数（Giga Floating-Point Operations Per Second，GFLOPs）、帧率（f/s）以及模型尺寸。AP用以度量实际检测中精度（P）与召回率（R）的综合表现，其计算公式如下：

$$AP = \int_0^1 P(R)\mathrm{d}R \quad (1)$$

由于目标检测中常用IoU度量预测框与真实框的重叠程度，并以此来划分预测结果的正负性。而不同的IoU阈值会产生不同的P-R曲线进而影响最终的AP值，因此为综合全面地反应模型性能，使用10个不同IoU阈值（0.50线性递增至0.95，间隔为0.05）对应的P-R曲线面积的平均值来表示AP值。

为评估不同网络的玉米籽粒计数性能，对180幅玉米测试集图像检测结果进行人工籽粒识别统计试验与评价。选用检测准确率（Detection Accuracy，DA）、漏检率（Miss-Detection Rate，MDR）、错检率（False-Detection Rate，FDR）、查准率（Precision Detection Rate，PDR）以及F_1值作为评价计数效果的指标，其计算公式如下。

$$DA = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$MDR = \frac{FN}{TP + FN} \times 100\% \quad (3)$$

$$FDR = \frac{FP}{TP + FP} \times 100\% \quad (4)$$

$$PDR = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$F_1 = \frac{2 \times DA \times PDR}{DR + PDR} \times 100\% \quad (6)$$

其中，TP为正确检测的玉米籽粒数，个；FP为错检的玉米籽粒数，个；FN为漏检的玉米籽粒数，个。

2.3 实验配置

图像处理硬件平台配置为：Intel（R） Core（TM） i5-10400F CPU @2.90 GHz处理器，16 G内存，1 TB硬盘，11 GB NVIDIA GeForce GTX 2080Ti GPU。编程软件为Pycharm2020社区版。深度学习框架为Pytorch 1.6，并使用CUDA11.3和cuDNN 8.2.0作为深度神经网络的并行计算架构和GPU加速库。

3　结果与讨论

3.1 网络模型训练

在同一实验配置条件下，采用2.2.1小节构建的玉米籽粒目标检测数据集对所选Mask R-CNN、EfficientDet-D5、YOLOv5-L、YOLOX-L四种网络进行训练与测试。模型训练过程中的损失值变化曲线如图8所示。由图8可知，各网络模型对玉米籽粒数据集均具有较强的拟合与泛化能力，且损失值的变化趋势基本相同。即在训练初期损失快速下降，中期震荡小幅减小，后期基本趋于稳定，网络模型实现收敛。模型的快速收

敛得益于Adam优化器的使用，其具有计算高效、自主调整学习率以及不受梯度伸缩变化影响的优点。

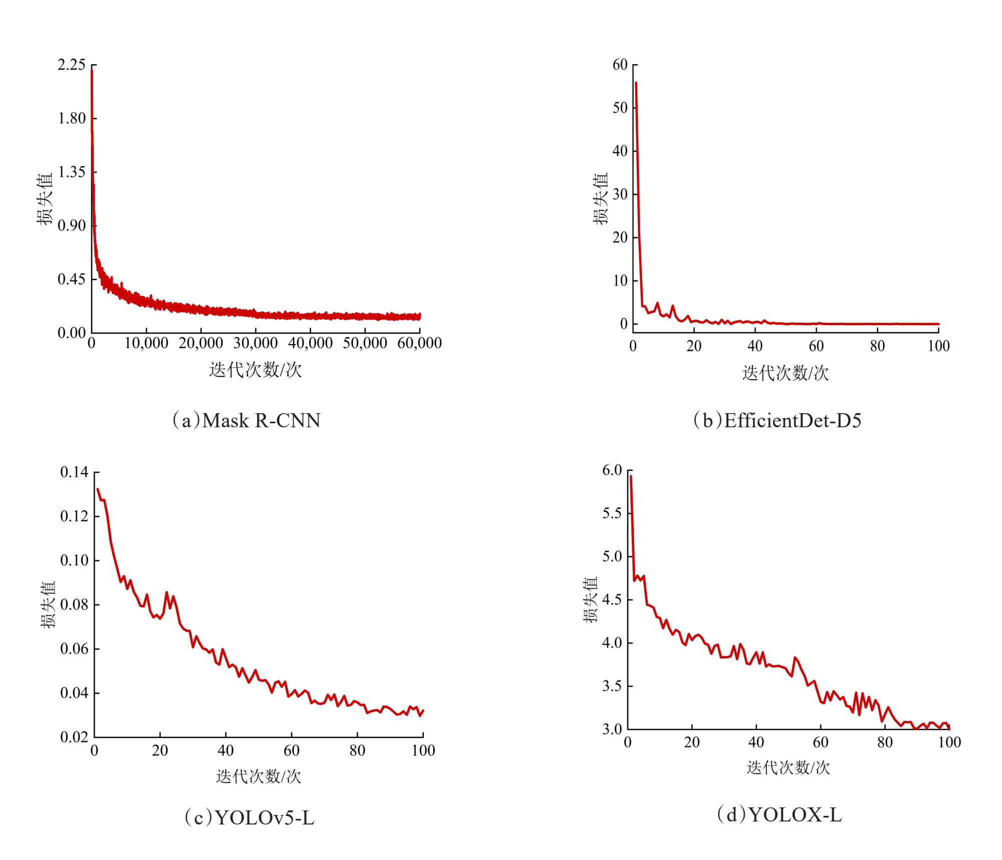


图8 四种目标检测网络的训练损失曲线

Fig. 8 Training loss curve of four object detection networks

经过对四种网络模型拟合泛化能力确认后，可通过对比模型在测试集上的整体检测性能来得到初步的最优模型。表1显示了四种网络的性能对比结果。整体而言，YOLOv5-L的各项指标值均取得最优，其AP值为78.3%，帧率为55.55 f/s，相较于YOLOX-L、Mask R-CNN和EfficientDet-D5分别提升了11.8（28.53）、9.5（46.7）和27.4（48.12）个百分点（帧数），这表明YOLOv5-L能够更快地提供更准确的检测结果。网络自身的结构差异导致对玉米图像关注度及特征学习能力的不同，是造成各网络平均精度存在差异性的主要原因。而相较于EfficientDet-D5和YOLOX-L，Mask R-CNN的AP值取得最优可能与其先生成候选框后回归分类的两阶段检测架构有关，但两阶段精度提升带来的计算资源的消耗使得其识别速率远低于直接进行回归分析的单阶段网络。当考虑模型复杂度及移动端部署可行性时，YOLOv5-L也表现出了极大的潜力，其GFLOPs、模型尺寸以及训练时长分别为最大值Mask R-CNN的37%、18.6%和5.1%，远优于上述性能依次递减的EfficientDet-D5和YOLOX-L。由于网络在移动端的部署限制可通过扩展内存等多种方法解决，但模型检测效果直接影响最终的籽粒计数精度与效率。结合上述分析可初步判定，四种网络用于籽粒检测的优劣顺序为：YOLOv5-L、Mask R-CNN、YOLOX-L和EfficientDet-D5。

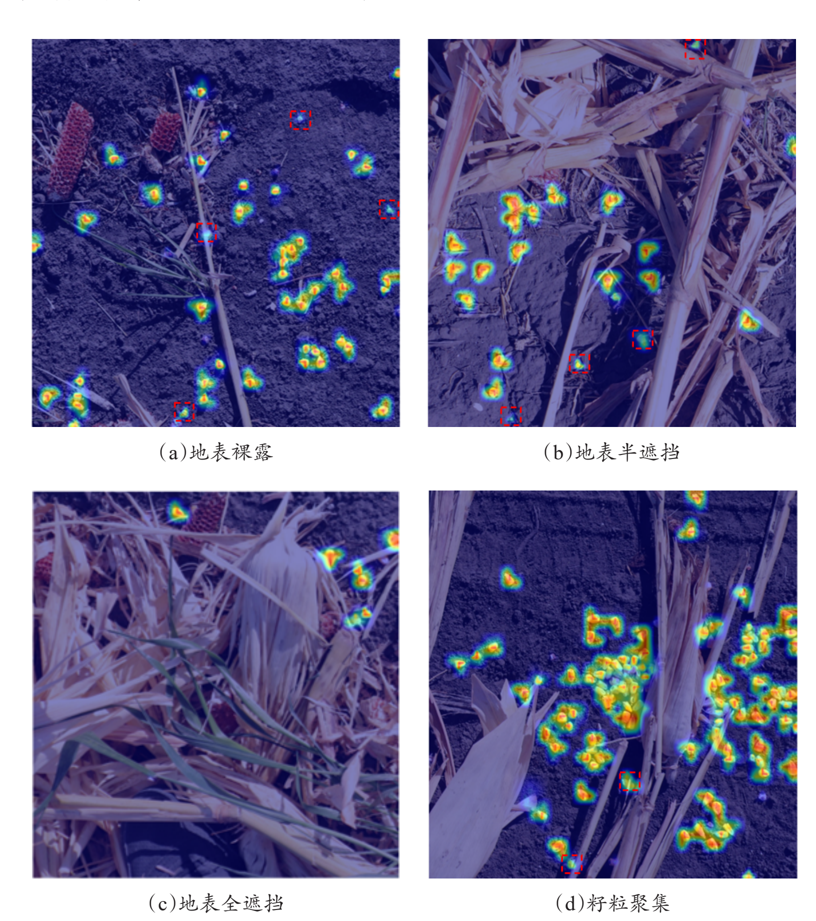
为了解输入图像的哪些区域使得最优模型YOLOv5-L做出了最终的识别决策，采用类激活热力图的方式对随机选取的四幅图像各区域的作用权重进行可视化（图9）。不同颜色代表不同的

表 1　四种网络模型性能测试结果对比

Table 1 Performance comparison of test results for four deep learning models

模型	AP/%	帧率/(f·s^{-1})	模型参数量/MB	GFLOPs	训练时长/min	模型尺寸/MB
Mask R-CNN	68.8	8.85	63.0	164.4	504	479.8
EfficientDet-D5	50.9	7.43	33.4	135.8	360	128.7
YOLOv5-L	78.3	55.55	47.0	115.6	26	89.3
YOLOX-L	66.5	27.02	54.0	155.3	84	207.0

权值，颜色越深权值越大。由图9可知，不同场景下，模型均是通过抑制背景区域，重点学习籽粒区域相关信息来实现识别决策的。在籽粒区域中又以玉米的胚及其胚轴区域对类别决策的贡献权值最大。进一步分析可知，相较于红色矩形框标注的干瘪、遮挡严重、破碎以及阴影处的玉米籽粒，色泽充盈、颗粒饱满的玉米籽粒对类别决策的贡献率更高，权值的偏狭可能导致权值较小的籽粒会被漏检或误检。



(a)地表裸露　(b)地表半遮挡

(c)地表全遮挡　(d)籽粒聚集

图 9　YOLOv5-L 模型的类激活热力图

Fig. 9 Class activation heatmap of YOLOv5-L model

3.2 识别结果分析

基于最优网络 YOLOv5-L 对测试集图像进行推理，考察复杂图像背景中玉米籽粒目标识别的效果，为避免籽粒标签影响后续识别结果的观测与统计，对标签进行了去除，部分示例如图 10 所示。由图 10 可知，YOLOv5-L 模型几乎能正确定位出不同场景下的所有玉米籽粒，当视野中图

像清晰度、光照以及地表状态发生变化时也能被很好地识别出来，表明构建的网络能稳定、可靠地检测玉米籽粒，对外部环境的干扰具有较强的鲁棒性。但对于不同的场景，模型也表现出了不同的性能。地表裸露和地表半遮挡场景的识别准确率要高于其余场景。相较于后两者，前两者的微弱损失主要以干瘪籽粒的误检为主，这可能与其籽粒离散且遮挡物较少使得模型过度学习籽粒特征造成过拟合有关，并从侧面证实了3.1节中干瘪籽粒区域对类别决策权值较小易造成误检的结论。对于后两者而言，其损失主要以籽粒堆叠或严重遮挡造成的漏检以及单框内包含两个籽粒的误检为主，漏检原因可能是对类别决策更重要的玉米胚及其胚轴区域被遮盖使得模型失去关键特征信息。而误检可能是采用非极大值抑制算法过滤置信度较低的边框时阈值设置不合理造成。进一步分析发现，网络的可靠性可能与籽粒颜色、形状等特征及其自身与周围区域的强烈过渡对比使得模型更易为目标识别提供更多有用信息有关。



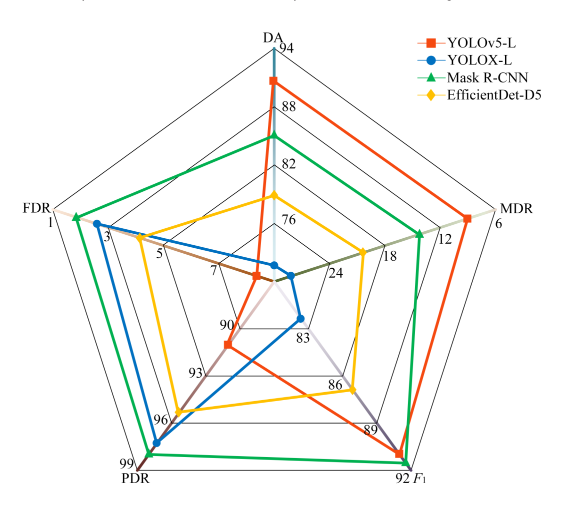
图10 YOLOv5-L模型对不同场景中玉米籽粒的识别结果示例

Fig. 10 Example of the detection results of YOLOv5-L model for corn kernels in different scenes

3.3 计数结果对比

为验证YOLOv5-L网络在籽粒计数中的可靠性和稳定性，采用检测准确率、漏检率、错检率、查准率和F_1值等指标对测试集进行整体评估，并与YOLOX-L、Mask R-CNN和Efficient-Det-D5三种网络进行比较，对比结果如图11所示。由图11可知，四种网络的检测性能各有不同。对于检测正确率和漏检率而言，YOLOv5-L的指标值最优分别为90.7%和9.3%，优于指标值

依次递减的Mask R-CNN、EfficientDet-D5、YOLOX-L三种网络5～19个百分点。而考虑错检率和查准率时，该排序得到逆转，即YOLOv5-L表现最差，并低于最优的Mask R-CNN网络7个百分点。在综合指标F_1中，该排序仅互换了YOLOv5-L和Mask R-CNN的位置，且两者的指标值分别为91.1%和91.6%，相差0.5%。明显地，较优的YOLOv5-L和Mask R-CNN网络的侧重优势不同。而实际生产中，玉米籽粒损失的计数应遵循宁错勿漏的原则，以避免因高估联合收割机性能而造成更大的收获损失。YOLOv5-L在查准率方面的劣势可能是由于瘪籽粒的大量错检引起的，本研究虽不对其进行计数，但干瘪籽粒的检出为后续进一步提高损失产量的估计精度提供了可能。此外，YOLOv5-L在检测效率、模型复杂度及应用潜力等方面的综合优势也为解决实际生产中高效便捷的核心需求，为加速研究商业化提供了便利。因此，结合上述分析可知，四种网络籽粒计数的优劣顺序依次为YOLOv5-L、Mask R-CNN、EfficientDet-D5和YOLOX-L。



注：DA、MDR、FDR、PDR、F_1分别指计数的检测准确率、漏检率、错检率、查准率以及检测准确率和查准率的调和平均值

图11 不同网络的籽粒计数性能对比

Fig. 11 Comparison of kernel counting performance of different networks

为明确各网络对不同场景类型的籽粒计数性能，对任一网络与场景组合的计数指标进行统计（表2）。明显地，YOLOv5-L和Mask R-CNN网络在不同场景下的DA、MDR和F_1值均高于其余网络，且两者中又以YOLOv5-L的DA和MDR为最优。而YOLOv5-L远低于其余三种网络的PDR和FDR，是造成不同场景下（除籽粒聚集外）其F_1值均略低于Mask R-CNN的主要原因。整体而言，不同场景下YOLOv5-L和Mask R-CNN的计数性能均较佳，且YOLOv5-L在含有籽粒较多的地表裸露、地表半遮挡、籽粒聚集三种场景下的高DA是其最终排序优于Mask R-CNN的主要原因。此外，四种网络均呈现出地表全遮挡和籽粒聚集的检测性能远低于其余场景的现象为下一步的研究提供了方向。

表2 不同场景下四种网络的玉米籽粒计数性能结果

Table 2 Corn kernel counting performance results of four networks in different scenarios

场景	指标	模型			
		Mask R-CNN	Efficient-Det-D5	YOLOv5-L	YOLOX-L
地表裸露	DA	95.6	88.9	98.2	86.3
	MDR	4.4	11.1	1.8	13.7
	FDR	1.7	3.8	8.6	3.2
	PDR	98.3	96.2	91.4	96.8
	F_1	96.9	92.4	94.7	91.2
地表半遮挡	DA	92.3	83.4	95.5	81.7
	MDR	7.7	16.6	4.5	18.3
	FDR	1.7	3.4	7.7	2.4
	PDR	98.3	96.6	92.3	97.6
	F_1	95.3	89.5	93.8	88.9
地表全遮挡	DA	84.0	56.4	76.1	61.5
	MDR	16.0	43.6	23.9	38.5
	FDR	5.7	7.5	9.3	2.9
	PDR	94.3	92.5	90.7	97.1
	F_1	88.9	70.1	82.8	75.3
籽粒聚集	DA	74.0	72.0	83.3	54.7
	MDR	26.0	28.0	16.7	45.3
	FDR	1.3	4.8	9.0	1.8
	PDR	98.7	95.2	91.0	98.2
	F_1	84.5	82.0	87.0	70.3

4 结 论

本研究以田间玉米籽粒检测为切入点，依据构建的不同目标检测网络对籽粒的识别，实现了收获过程中玉米籽粒损失计数，验证了深度学习

技术在该项任务中的可行性。结果表明，四种模型中，YOLOv5-L的性能最佳，其计数的DA、MDR分别为90.7%和9.3%，计数效果优于Mask R-CNN、EfficientDet-D5和YOLOX-L网络，处理速度为55.55 f/s，可满足籽粒损失的实时监测与收割质量的快速评估，因此可作为玉米联合收割机收获损失精准控制信息系统和检测装置开发的核心算法。

研究仍存在一定不足，未来的工作将从以下几方面展开：（1）针对地表全遮挡和籽粒聚集场景较高的漏检率问题，拟引入注意力和特征增强机制对模型进行改进，以进一步提高计数精度；（2）探究籽粒色泽程度、干瘪程度以及自身大小等变化对识别的影响，开发检测结果的二次判别算法以实现错检率的降低；（3）引入更丰富的数据类别，如不同模糊度、亮度、角度、玉米品种等，并扩充训练图像集，以提高模型的鲁棒性与稳定性。最终应重视收割过程中的灰尘问题，并开发相应的遮蔽装置，以实现收割过程中损失的实时检测。

参考文献：

[1] 张伟锋, 刘新娇, 李秋枫, 等. 世界玉米生产与贸易概况分析[J]. 世界农业, 2014(3): 111-114.
ZHANG W, LIU X, LI Q, et al. General situation analysis of world corn production and trade[J]. World Agriculture, 2014(3): 111-114.

[2] WU Y, LI X, MAO E, et al. Design and development of monitoring device for corn grain cleaning loss based on piezoelectric effect[J]. Computers and Electronics in Agriculture, 2020, 179(12): ID 105793.

[3] XU L, WEI C, LIANG Z, et al. Development of rapeseed cleaning loss monitoring system and experiments in a combine harvester[J]. Biosystems Engineering, 2019, 178: 118-130.

[4] VALIENTE-GONZÁLEZJM, ANDREU-GARCÍAG, POTTER P, et al. Automatic corn (*Zea mays*) kernel inspection system using novelty detection based on principal component analysis[J]. Biosystems Engineering, 2014, 117: 94-103.

[5] ORLANDI G, CALVINI R, FOCA G, et al. Automated quantification of defective maize kernels by means of multivariate image analysis[J]. Food Control, 2018, 85: 259-268.

[6] LI X, DAI B, HONG S, et al. Corn classification system based on computer vision[J]. Symmetry, 2019, 11 (4): 591-591.

[7] 张恬, 赵德安, 周童. 图像处理在联合收割机夹带损失检测中的应用[J]. 农机化研究, 2009, 31(4): 70-72.
ZHANG T, ZHAO D, ZHOU T. Application of image processing on combine harvester attachment loss[J]. Journal of Agricultural Mechanization Research, 2009, 31(4): 70-72.

[8] 辛博, 吴涛, 陈春林, 等. 一种基于图像处理的谷物收割清选损失实时在线检测方法: CN107123115A[P]. 2017-09-01.
XIN B, WU T, CHEN C, et al. A real-time online detection method for grain harvesting and cleaning loss based on image processing: CN107123115A[P]. 2017-09-01.

[9] WELLINGTON C K, BRUNS A J, SIERRA V S, et al. Grain quality monitoring: US10664726B2[P]. 2017-10-02.

[10] 孙红, 李松, 李民赞, 等. 农业信息成像感知与深度学习应用研究进展[J]. 农业机械学报, 2020, 51(5): 1-17.
SUN H, LI S, LI M, et al. Research progress of image sensing and deep learning in agriculture[J]. Transactions of the CSAM, 2020, 51(5): 1-17.

[11] MONHOLLEN N S, SHINNERS K J, FRIEDE J C, et al. In-field machine vision system for identifying corn kernel losses[J]. Computers and Electronics in Agriculture, 2020, 174: ID 105496.

[12] 吕璐, 程虎, 朱鸿泰, 等. 基于深度学习的目标检测研究与应用综述[J]. 电子与封装, 2022, 22(1): 72-80.
LYU L, CHENG H, ZHU H, et al. Progress of research and application of object detection based on deep learning[J]. Electronics & Packaging, 2022, 22(1): 72-80.

[13] 董丽君, 曾志高, 易胜秋, 等. 基于YOLOv5的遥感图像目标检测[J]. 湖南工业大学学报, 2022, 36(3): 44-50.
DONG L, ZENG Z, YI S, et al. Research on a YOLOv5-Based remote sensing image target detection[J]. Journal of Hunan University of Technology, 2022, 36 (3): 44-50.

[14] 谢富, 朱定局. 深度学习目标检测方法综述[J]. 计算机系统应用, 2022, 31(2): 1-12.
XIE F, ZHU D. Survey on deep learning object detection[J]. Computer Systems & Applications, 2022, 31 (2): 1-12.

[15] 寇大磊, 权冀川, 张仲伟. 基于深度学习的目标检测框架进展研究[J]. 计算机工程与应用, 2019, 55(11): 25-34.
KOU D, QUAN J, ZHANG Z. Research on progress of object detection framework based on deep learning[J]. Computer Engineering and Applications, 2019, 55(11): 25-34.

[16] 包晓敏, 王思琪. 基于深度学习的目标检测算法综

述[J]. 传感器与微系统, 2022, 41(4): 5-9.
BAO X, WANG S. Survey of object detection algorithm based on deep learning[J]. Transducer and Microsystem Technologies, 2022, 41(4): 5-9.
[17] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]// The IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2017: 2961-2969.
[18] TAN M, PANG R, LE Q. Efficientdet: Scalable and efficient object detection[C]// The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2020: 10781-10790.
[19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. arXiv:2004.10934[cs.CV], 2020.
[20] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding yolo series in 2021[J/OL]. arXiv:2107.08430, 2021.

Infield Corn Kernel Detection and Counting Based on Multiple Deep Learning Networks

LIU Xiaohang[1,2], ZHANG Zhao[1,2*], LIU Jiaying[1,2], ZHANG Man[1,2], LI Han[1,2], Paulo FLORES[3], HAN Xiongzhe[4,5]

(1. *Key Laboratory of Smart Agriculture System Integration*, *Ministry of Education*, *China Agricultural University*, *Beijing* 100083, *China*; 2. *Key Lab of Agricultural Information Acquisition Technology*, *Ministry of Agriculture and Rural Affairs*, *China Agricultural University*, *Beijing* 100083, *China*; 3. *Department of Agricultural and Biosystems Engineering*, *North Dakota State University*, *Fargo*, *ND* 58012, *United States*; 4. *Department of Biosystems Engineering*, *Kangwon National University*, *Chuncheon* 24341, *Korea*; 5. *Interdisciplinary Program in Smart Agriculture*, *Kangwon National University*, *Chuncheon* 24341, *Korea*)

Abstract: Machine vision has been increasingly used for agricultural sensing tasks. The detection method based on deep learning for infield corn kernels can improve the detection accuracy. In order to obtain the number of lost corn kernels quickly and accurately after the corn harvest, and evaluate the corn harvest combine performance on grain loss, the method of directly using deep learning technology to count corn kernels in the field was developed and evaluated. Firstly, an RGB camera was used to collect image with different backgrounds and illuminations, and the datasets were generated. Secondly, different target detection networks for kernel recognition were constructed, including Mask R-CNN, EfficientDet-D5, YOLOv5-L and YOLOX-L, and the collected 420 effective images were used to train, verify and test each model. The number of images in train, verify and test datasets were 200, 40 and 180, respectively. Finally, the counting performances of different models were evaluated and compared according to the recognition results of test set images. The experimental results showed that among the four models, YOLOv5-L had overall the best performance, and could reliably identify corn kernels under different scenes and light conditions. The average precision (AP) value of the model for the image detection of the test set was 78.3%, and the size of the model was 89.3 MB. The correct rate of kernel count detection in four scenes of non-occlusion, surface mid-level-occlusion, surface severe-occlusion and aggregation were 98.2%, 95.5%, 76.1% and 83.3%, respectively, and F_1 values were 94.7%, 93.8%, 82.8% and 87%, respectively. The overall detection correct rate and F_1 value of the test set were 90.7% and 91.1%, respectively. The frame rate was 55.55 f/s, and the detection and counting performance were better than Mask R-CNN, EfficientDet-D5 and YOLOX-L networks. The detection accuracy was improved by about 5% compared with the second best performance of Mask R-CNN. With good precision, high throughput, and proven generalization, YOLOv5-L can realize real-time monitoring of corn harvest loss in practical operation.

Key words: harvest loss; infield corn kernel; deep learning; kernel count; YOLOv5-L; YOLOX-L; Mask R-CNN; EfficientDet-D5